

## REMARKS

Claims 1-24 are rejected. Claims 1-24 remain pending. Claims 1,4, 6-9, 12, 14-17, 20, and 22-24 are amended herein. No new matter is introduced as a result of the Claim amendments.

### Allowable Subject Matter

The Applicants wish to thank the Examiner for indicating the allowable subject matter of Claims 6-8, 14-16, and 22-24.

### 35 U.S.C. § 103 Rejections

Claims 1-5, 9-13, and 17-21 are rejected under 35 U.S.C. 103(a) as being anticipated by Thompson et al. (US 2002/0103841A1), hereinafter referred to as "Thompson." The Applicants respectfully submit that the embodiments of the present invention recited in Claims 1-24 are not taught or suggested by Thompson.

Claim 1 of the present invention recites (emphasis added):

- a) generating a list of reference words and phrases and a list of non-reference words and phrases from a selected group of documents;
- b) comparing said list of reference words and phrases with a joined list containing said reference words and phrases and said non-reference words and phrases, using an edit-distance algorithm to create an approximate duplicates list;
- c) filtering said approximate duplicates list to create a thesaurus of standard words and phrases and their variations; and
- d) editing said selected group of documents with an editor operable to use said thesaurus to replace a word or phrase on said approximate duplicates list with said standard words and phrases.

Claims 9 and 17 recite similar claim limitations. The Applicants respectfully submit that Thompson does not teach or suggest that the list of reference words comes from the set of selected documents. Instead, Thompson relies upon one or more of the user-selected reference dictionaries to determine which words or phrases populate the list of non-reference words. The Applicants respectfully submit that this teaches away from the recited limitation that a list of reference words is generated from a selected group of documents. The "reference words" used by Thompson are supplied by whichever previously created dictionary the user chooses. In contrast, the present invention recites populating both the reference word list and the non-reference word list with words and phrases which come from the selected group of documents being examined. Thus, the Applicants respectfully submit that the cited reference teaches away from the recited claim limitation of generating a list of reference words and phrases from a selected group of documents.

The Applicants further submit that Thompson does not teach or suggest the recited claim limitation of (emphasis added):

comparing said list of reference words and phrases with a joined list containing said reference words and phrases and said non-reference words and phrases, using an edit-distance algorithm to create an approximate duplicates list.

The cited reference does not teach or suggest joining the list of reference words and phrases with the list of non-reference words and phrases, nor does it teach or suggest comparing the original reference list of words and phrases with the joined list as recited in the present invention. Thus, the Applicants respectfully submit that the recited claim limitations of Claims 1, 9, and 17 of the present invention are not rendered obvious by the teaching of Thompson. Accordingly, the Applicants respectfully submit that the rejections of Claims 1, 9, and 17 of the present invention under 35 U.S.C. § 103(a) are overcome.

With reference to Claims 2, 10, and 18 of the present invention, the Applicants respectfully submit that Thompson does not teach or suggest the recited claim limitation of discarding words and phrases from said selected group of documents that are on a stop word. The rejection cites paragraph 0492 as disclosing the claimed limitation. However, the Applicants respectfully submit that Thompson merely teaches that if a document comprises a given percentage of unrecognized words, it is not processed. As discussed on page 8 of the present invention, stop words are defined as words not regarded as relevant to the current domain. This may include commonly used words, numbers, or characters which are not considered critical for the purposes of normalization. However, the present invention does not recite a cessation of the processing of the document if the percentage of stop words exceeds a given percentage as taught by the cited reference. Accordingly, the Applicants respectfully

submit that the rejections of Claims 2, 10, and 18 of the present invention under 35 U.S.C. § 103(a) are overcome.

With reference to Claims 3, 11, and 19 of the present invention, the Applicants respectfully submit that Thompson does not teach or suggest the recited claim limitation that the words and phrases not discarded comprise the lists of reference and non-reference words and phrases. As discussed above with reference to Claim 1, the Applicants respectfully submit that Thompson clearly teaches that the words and phrases populating the reference word list come from previously created dictionaries and not from the selected documents being processed as claimed in Claims 3, 11, and 19. Accordingly, the Applicants respectfully submit that the rejections of Claims 3, 11, and 19 of the present invention under 35 U.S.C. § 103(a) are overcome.

With reference to Claims 4, 12, and 20 of the present invention, the Applicants respectfully submit that Thompson does not teach or suggest the recited claim limitations of (emphasis added):

- a1) counting the frequency of occurrence of a plurality of words and phrases from said selected group of documents;
- a2) placing words and phrases with special characters embedded within them on said reference word list;
- a3) processing words and phrases from said selected group of documents not already on said reference word list with a spell-checker program, wherein words and phrases that are recognized as correctly spelled are placed on said reference word list and all unrecognized words and phrases are placed on said non-reference word list;

a4) setting a frequency of occurrence threshold for said reference word list, wherein words and phrases which have a frequency of occurrence below said threshold are discarded as irrelevant; and

a5) setting a word frequency threshold for said non-reference word list, wherein words and phrases which have a frequency of occurrence above said threshold remain on said non-reference word list.

The Applicants respectfully submit that the cited passages of Thompson do not teach or suggest the recited limitations of Claims 4, 12, and 20 shown above. Instead, Thompson teaches that the document is rated in its entirety as to the quality of the document based upon the number of valid/non-valid terms. However, Thompson fails to teach or suggest placing words and phrases with special characters embedded within them on the reference word list. In fact, Thompson teaches away from the recited limitation in paragraph 0513 which teaches that words with embedded characters are regarded as errors which are automatically put on the "non-reference" list.

The Applicants further submit that Thompson does not teach or suggest discarding words or phrases from the reference word list having a frequency below a threshold. In contrast, because the reference word list of Thompson comprises previously created reference dictionaries, it would be counter-intuitive to discard words from them. Similarly, Thompson does not teach or suggest setting a word frequency threshold for the non-reference word list, wherein words and phrases which have a frequency of occurrence above the threshold remain on said non-reference word

list. Accordingly, the Applicants respectfully submit that the rejections of Claims 4, 12, and 20 of the present invention under 35 U.S.C. § 103(a) are overcome.

## CONCLUSION

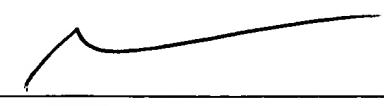
Based on the arguments presented above, the Applicants respectfully assert that Claims 1-24 overcome the rejections of record and, therefore, the Applicants respectfully solicit allowance of these Claims.

The Applicants have reviewed the references cited but not relied upon. The Applicants did not find these references to show or suggest the present claimed invention: U.S. 6,353,840, U.S. 6,687,873.

The Examiner is invited to contact Applicants' undersigned representative if the Examiner believes such action would expedite resolution of the present Application.

Respectfully submitted,  
WAGNER, MURABITO & HAO LLP

Date: 8/9/05

  
John P. Wagner, Jr.  
Reg. No. 35,398

Two North Market Street  
Third Floor  
San Jose, California 95113  
(408) 938-9060